Multi-dimensional Kernel smoothing:

$$\hat{\mu}(x_0) = \sum_{i=1}^{n}\left(Y_i \frac{\prod_{j=1}^{p} K\left(\frac{x_{0j} - x_{ij}}{h_j}\right)}{\sum_{k=1}^{n}\left[\prod_{j=1}^{p} K\left(\frac{x_{0j} - x_{kj}}{h_j}\right)\right]}\right)$$

- As $n \to \infty$ for kernel regression, with the optimal bandwidth, the bias and variance of the model approach 0

$$MSE = \sigma^2 + O(h_{opt}^4) = \sigma^2 + O(n^{\frac{-4}{4+p}})$$

- Kernel regression converges slowly
- KNN, splines, trees, random forests, deep neural nets all are universally consistent, but all of these converge to μ at the same rate (called mini-max rate)

Additive Model

$$Y = \alpha + \sum_{i=1}^{p} f_i(X_i) + \epsilon$$ (each X_i gets its own function)

- f_i are called partial response functions

How to Fit:

$$f_j(x_j) + \epsilon = Y - (\alpha + \sum_{k \neq j} f_k(x_k))$$

1. start with $\alpha = \bar{y}$, $f_1 = f_2 = \dots = f_p = 0$
2. For $j \in 1:p$
   a) get residual for $y_i - \left(\alpha + \sum_{k \neq j} f_k(x_k)\right)$
   b) use linear smoother to regress on residuals (say we get function s_j)
   c) $f_j = s_j - \bar{s}_j$
3. repeat loop until things stop changing

Spline:

$$S_\lambda = \underset{m}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (y_i - m(x_i))^2 + \lambda \int \left(\frac{d^2 m}{dx^2}(x)\right)^2 dx$$

- computationally faster to fit spline than kernel

Hypothesis Testing with MSE

- Suppose we want to test if $\mu(x)$ is a linear function

H_0: $\mu(x)$ is a linear function H_a: $\mu(x)$ is not linear

Test statistic: D = difference in MSE

- If linear model is right it will predict at least as well as with non-parametric model
- If linear model is wrong, eventually non-parametric model will beat linear model